Midterm Project

Participate in the *House Prices - Advanced Regression Techniques* competition in *kaggle.com*.

Link: https://www.kaggle.com/c/house-prices-advanced-regression-techniques

Follow the below steps to submit your first prediction:

1. Create user id and password in kaggle.com
2. Join the *House Prices - Advanced Regression Techniques* competition
3. Download the train and test data
4. Merge the two datasets
5. Substitute missing values with appropriate values
6. Substitute outliers with appropriate values
7. Perform exploratory data analysis using visualizations to identify basic patterns
8. Use *statsbuddy.net* & *R Studio Cloud* to predict the *SalePrice* column in test data
9. Manually delete all columns, other than *Id* and *SalePrice*, from the prediction file
10. Submit your prediction in *kaggle.com* and get your score and rank

Try to improve your score and rank by trying:

- Different sets of independent variables, depending on their level of influence
- Different machine learning models (e.g., Linear Regression, Decision Tree, and Random Forest, SVM, XGBoost etc.)
- Different hyper-parameters (e.g., number of levels, and number of trees, as applicable)

You might consider the following to improve your score/rank:

- Create new meaningful columns
- Convert numeric variables to categorical variables
- Convert categorical variables to numeric variables

Deliverables:

Write a two-page report describing your approach and journey. Describe what you tried (and what you didn't try) and why (very important!). No R codes are required. Submit the report and a screenshot of the leaderboard from *kaggle.com* that shows your score and rank.

Distribution of Points:

10 points for your score and rank, and 15 points for the analysis and report. The first 10 points will be distributed based on your relative rank in the class (i.e., 10 for the top 10% students in terms of rank, 9 for the next 20% students, 8 for the next 30% students and 7 for the remaining students.)